BRINGING PATIENT POPULATIONS TO THE INTEGROME

Kohane, Isaac S.[1], Butte, Atul[2]; Valtchinov, Vlad[3]; Murphy, Shawn N.[1]
[1]Harvard Medical School, Boston, MA; [2]Stanford School of Medicine, Stanford, CA;
[3]Partners Healthcare Systems, Boston, MA
Keywords: Integrome, populations, translational medicine, genomics

The Integrome project of I2B2 has as its goal the redefinition of human disease based on combined genomic and clinical data. Why is this important? The current classification of disease is a mix of attributing specific diseases to specific organ systems (e.g. heart disease) or to specific clinical manifestations (e.g., diabetes mellitus, named after the presence of glucose in the urine). These classifications have been the mainstay of medical education and practice for decades and even centuries. As a result, the commonalities across these diseases, the underlying pathophysiological processes that span these various diseases are only glimpsed at from the particular perspective of this historically accreted rather arbitrary classification scheme.

As an initial foray into developing a data-driven robust view of disease that includes all genomic data and clinical findings, *the Integrome*, i2b2 investigators obtained all the measures of genes in the National Library of Medicine's Gene Expression Omnibus, a public database and used artificial intelligence techniques to read the textual descriptions of these tens of thousands of experiments and assign these experiments automatically to one or more disease or process categories (e.g., heart failure or aging). In parallel, the investigators took the tens of thousands of genes that were measured in each of these experiments and determined through millions of calculations which genes were truly characteristic of the process or disease described in that experiment and those that were not. This result provided several interesting insights. For example diseases such as gastritis had gene signatures remarkably similar to those of heart attacks, and several genes widely known to be associated with inflammation were associated with a very large array of diseases.

In addition to integrating genomic measures to disease categories, we are now going to integrate further the clinical manifestations of the 2.5 million patients in the Partners Health Care system electronic medical record (that has been extracted, with due protection of patient privacy) into the i2b2 clinical research chart. This next step will not only allow us to understand how we can reclassify diseases based on their genomic signatures but will allow us to determine how individual patients can be recategorized based on their relationship to one another within the Integrome. In doing so several important challenges will have been overcome that all similar investigations must address.

E-mail: Isaac_kohane@harvard.edu